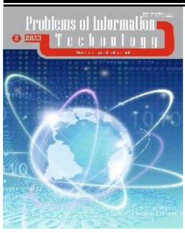

Available online at www.jpit.az

14 (2)
2023

Comparative analysis of models for solar station output prediction

Javad V. Najafli

Azerbaijan Technical University, H.Javid avenue, 25, AZ1073 Baku, Azerbaijan

qwerty642.5@mail.ru

ARTICLE INFO

http://doi.org/10.25045/jpit.v14.i2.04

Article history:
Received 24 February 2023
Received in revised form 31 March 2023
Accepted 18 May 2023

Keywords:
Solar energy
Artifical Intelligence
Forecasting
Solar power generation
Energy sustainability

ABSTRACT

This research paper explores the prediction of solar energy radiation using various machine learning methods and neural networks. The results are presented based on the analysis of four different datasets obtained from solar stations. The study begins with an overview of solar energy in the context of contemporary challenges in the fields of energy and environmental sustainability, and reviews previous research related to the application of artificial intelligence in solar energy. The main contribution of the work lies in the analysis and comparison of diverse machine learning models and neural networks for predicting solar energy radiation. The results are compared considering accuracy metrics (RMSE - Root Mean Squared Error, MAE - Mean Absolute Error, MRE - Mean Relative Error) and execution times for each model. Each model is evaluated on four datasets with different characteristics.

1. Introduction

In recent decades, as energy demands have risen and environmental issues associated with traditional energy sources have escalated, global attention has shifted towards seeking sustainable and environmentally friendly alternatives. In this context, solar energy has emerged as a key focus of research and innovation in the field of energy, aiming to address these challenges. Various researchers, such as Ibrahimov (2010), Aydin and Ulviyye (2019), as well as Vidadili et al (2017), have extensively explored the potential of solar energy and its prospects, particularly in country rich in solar resources like Azerbaijan.

The undeniable significance of solar energy for Azerbaijan is driven not only by the pursuit of energy portfolio diversification but also by the recognition of the need to reduce greenhouse gas emissions and mitigate climate change. Harnessing energy from solar sources can not only improve the country's environmental situation but also ensure long-term energy stability by reducing dependence on imported fossil fuels.

Research indicates that Azerbaijan possesses substantial potential for effective solar energy utilization across its territory. Moreover, the works by Mayis, Elchin and Gulsura (2020), Feyruz (2021), and Bahman (2023) emphasize the necessity of developing integrated energy storage and management systems to ensure uninterrupted power supply in the face of variable solar activity.

This study focuses on researching and developing innovative methods that contribute to enhancing the efficiency of solar energy in Azerbaijan using artificial intelligence (AI). It aims to synergize the potential of solar energy with AI capabilities to optimize processes in this domain.

The structure of our work is designed to systematically investigate key aspects of this topic. It is commenced with an introduction that contextualizes the research, outlines our objectives and tasks.

Subsequently, this study reviews solar energy in Azerbaijan. This section analyzes the current state and significance of solar energy in the country. It explores the actual contribution of solar energy to the energy system and identify its potential in ensuring energy supply sustainability.

Advancing the research, the section reviews the application of artificial intelligence in solar energy. This section examines various methods and approaches that AI can offer to optimize the functioning of solar stations. The tasks that can be tackled using AI, such as energy production forecasting and equipment operation optimization are highlighted.

Then the proposed methodology is described. This section elaborates on the principles and algorithms developed to enhance solar energy efficiency through AI.

Continuing, the study presents the results of experiments and studies conducted in this field. We analyze data, compare results with existing approaches, and highlight the advantages of our method.

Concluding section summarizes the work, outlines key findings, draws conclusions, and discusses potential directions for future research in the realm of AI application in solar energy in Azerbaijan.

2. Related work

In recent years, numerous studies have been presented, covering various aspects of AI usage in solar energy.

The project by Ahmad et al. (2017) explored the application of neural networks for forecasting solar activity and energy generation based on data from solar panels. Their research employed deep neural networks to analyze and predict the variability of solar radiation, enhancing the efficiency of energy generation from solar panels.

The method of adaptive optimization of solar panel tilt angles using genetic optimization algorithms was introduced in the work by Mahdi Khodayar et al. (2019). The study demonstrated that employing genetic algorithms allows for determining optimal panel tilt angles based on time of day and geographical location, ultimately enhancing the effectiveness of solar power plants.

The utilization of machine learning methods for forecasting solar radiation intensity was investigated in the study by Liu and Kun (2019). The authors proposed a prediction model based on random forest algorithms, showcasing high prediction accuracy—a crucial factor for planning and managing solar power plants.

The research by Hoofar Hemmetabady et al. (2022) focused on optimizing the distribution of energy from solar panels within a grid. The authors proposed a hybrid system that combines AI and grid management technologies, effectively distributing and storing generated energy as per demand.

These studies illustrate the diversity of approaches in using artificial intelligence within solar energy, highlighting its potential to enhance the efficiency, reliability, and sustainability of solar power plants.

This article conducts a comprehensive study of various forecasting methods, including Linear Regression, Random Forest, XGBoost, CatBoost, GradientBoost, as well as deep learning methods such as GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory). An approach here facilitates effective comparison of these methods based on key metrics like RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and MRE (Mean Relative Error) while considering elapsed time.

The study adapts and analyzes these methods specifically for solar station data from different regions of the world, resulting in more accurate forecasts and optimized solar energy utilization. The findings offer valuable recommendations for selecting the best forecasting method within the context of Azerbaijan's energy system.

Ultimately, this research contributes significantly to improving the efficiency and sustainability of solar energy in Azerbaijan, providing opportunities for radiation forecasting using advanced AI methods.

3. Proposed approach

This section provides a brief description and formulas for the methods employed in this research for forecasting the efficiency of solar stations:

3.1 Linear Regression: Linear regression builds a model of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$, where y is the target variable, $x_1, x_2, ..., x_k$ are the input features, and $\beta_0, \beta_1, \beta_2, ..., \beta_k$ are the coefficients adjusted during the learning process. The model tunes these coefficients to minimize the sum of squared errors (MSE).

3.2 Random Forest: Random Forest consists of multiple decision trees. Each tree is trained on a random subset of data. The prediction is obtained by averaging the predictions of all trees. The model's prediction for an object x is formulated as $F(x) = (\frac{1}{N}) \cdot \sum_i f_i(x)$, where N is the number of trees, $f_i(x)$ is the prediction of the i th tree.

3.3 XGBoost и CatBoost: These are gradient boosting methods. The model's prediction at each step is constructed as the sum of predictions from all previous trees, multiplied by learning rates. The error at each step is adjusted to improve predictions. The model's prediction for an object x is given by $F_k(x) = F_{k-1}(x) + \eta \cdot h_k(x)$, where $F_k(x)$ is the prediction at step k , η is the learning rate, and $h_k(x)$ is the prediction of the new tree at step k .

3.4 GradientBoost: The working principle is analogous to XGBoost and CatBoost. The model's prediction at each step is also constructed as the sum of predictions from all previous trees, multiplied by learning rates. The error at each step is adjusted to enhance predictions.

3.5 StochasticBoost: This is a variant of gradient boosting where, at each step, a random subset of data is chosen for training a new tree. This introduces randomness and diversity into the model.

3.6 GRU и LSTM: These are deep learning models for analyzing sequential data. They employ internal mechanisms to retain information about dependencies in time series data. Formulas for GRU and LSTM are more intricate, encompassing activation operations, weight coefficients, and internal states to account for data dependencies.

Our approach to the work is demonstrated in the flowchart (Fig. 1) below:

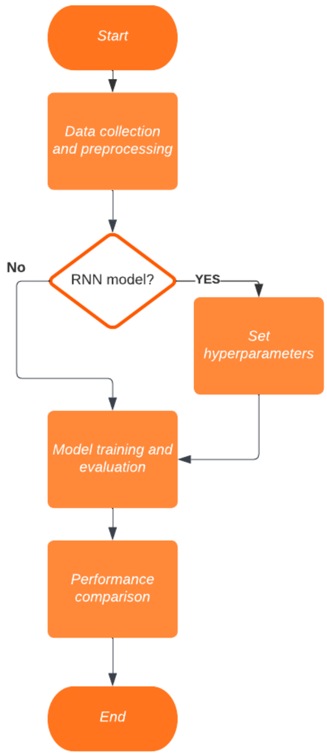

Fig.1. Process visualization

4. Experimental results

This study utilizes four distinct datasets obtained from open sources to validate the performance of the methods proposed here. Quality Metrics: To assess prediction quality, three main metrics are employed — RMSE, MAE, MRE. Lower values of RMSE, MAE, MRE indicate more accurate predictions. Each model is also evaluated in terms of execution time, expressed in seconds. The hyperparameters selected for the GRU and LSTM models are as follows: The number of neurons in the input layer is set to 256. The optimizer employed is nadam, and the loss function used is MSE. The training process is carried out for a total of 40 epochs, with early stopping implemented and a patience value of 4 epochs. The batch size for each dataset is set to 90, 29, 90, and 96, respectively.

Below are tables presenting the results for each dataset:

Table 1. Solar power Generation Dataset.

Model	RMSE	MAE	MRE	El.Time (sec)
Linear Reg.	0.1594	0.0255	10.7%	0.03
Random Forest	0.0835	0.007	2.8%	2.72
XGBoost	0.0844	0.0071	3%	0.53
CatBoost	0.0757	0.0057	2.8%	6.15

GradientBoost	0.0804	0.0065	3.1%	1.12
StochasticBoost	0.0801	0.0064	2.8%	1.75
GRU	0.1023	0.0549	25.4%	22.17
LSTM	0.1058	0.0536	26.3%	32.33

Table 2. Solar Power Plant Dataset.

Model	RMSE	MAE	RME	El.Time (sec)
Linear Reg.	0.11675	0.01366	5.2%	0.02
Random Forest	0.07457	0.00557	2.5%	6.57
XGBoost	0.07481	0.00563	2.6%	0.78
CatBoost	0.06905	0.00478	2.4%	8.42
GradientBoost	0.08125	0.00661	3.1%	2.47
StochasticBoost	0.07034	0.00497	2.3%	1.71
GRU	0.10778	0.06692	20.3%	71.43
LSTM	0.09878	0.05771	20%	142.73

Table 3. Solar Radiation Prediction Dataset.

Model	RMSE	MAE	MRE	El.Time (sec)
Linear Reg	0.1205	0.01452	7.8%	0.05
Random Forest	0.0497	0.00248	1.5%	40.07
XGBoost	0.05266	0.00278	1.9%	3.64
CatBoost	0.05251	0.00276	2%	14.06
GradientBoost	0.06689	0.00448	2.9%	11.2
StochasticBoost	0.05284	0.0028	1.9%	3.15
GRU	0.06952	0.04577	16%	104.4
LSTM	0.06562	0.03658	16.8%	245.49

Random Forest, XGBoost и CatBoost Random Forest, XGBoost, and CatBoost demonstrate favorable results in terms of both prediction accuracy (low RMSE and MAE values) and reasonable execution times. StochasticBoost also yields low error values and performs well in terms of execution speed across various datasets. GRU and LSTM (neural networks) deliver good results but involve significantly longer training and execution times compared to other models. Linear Regression provides less accurate predictions compared to more complex models.

In the final dataset, three variables are predicted simultaneously: Diffuse Horizontal Irradiance, Direct Normal Irradiance, and Global Horizontal Irradiance:

Table 4. Solar Radiation Dataset

Model	RMSE	MAE	MRE	Elapsed Time (sec)
Linear Regression	0.11909	0.01418	6.1%	0.2
Random Forest	0.07013	0.00492	2.1%	208.39
XGBoost	0.07229	0.00523	2.4%	46.32
GRU	0.09735	0.05143	6.4%	170.78
LSTM	0.10018	0.04698	5.9%	262.81

5. Conclusion and future work

This study investigated and compared various machine learning models and neural networks for forecasting solar radiation based on four distinct datasets. The obtained results allowed for identifying the strengths and weaknesses of each model, as well as determining the most suitable methods for specific requirements.

The research considered different aspects—prediction accuracy, execution time, and model versatility. Based on the provided data, one can select a model depending on specific needs. If precision is crucial and minor time costs are acceptable, more complex models like Random Forest, XGBoost, and CatBoost might be preferable. For a balanced approach between accuracy and execution time, StochasticBoost could be a good choice. In cases where neural networks are necessary and longer training is tolerable, GRU and LSTM also offer reasonable accuracy.

Even though linear regression had lower accuracy compared to other methods, it could be useful in situations where model simplicity and interpretability were crucial.

In the future, additional research focusing on hyperparameter optimization for each model could enhance their accuracy and efficiency. Exploring new feature extraction methods from data could also contribute to model prediction improvements. Combining predictions from different models or invecomstigating the impact of additional variables or factors on radiation, if available, could lead to improved forecasts.

Overall, this study proved valuable in identifying optimal models for predicting solar radiation on various datasets. Subsequent research could delve deeper into the selected models, taking into account the proposed directions for development.

References

Alzahrani, A., Shamsi, P., Dagli, C., & Ferdowsi, M. (2017). Solar irradiance forecasting using deep neural networks. Procedia Computer Science, 114, 304-313.

Aydin, U. (2019). Energy insecurity and renewable energy sources: Prospects and challenges for Azerbaijan.

Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25, 197-227.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

Culp, M., Johnson, K., & Michailides, G. (2007). ada: An r package for stochastic boosting. Journal of statistical software, 17, 1-27.

Gulaliyev, M. G., Mustafayev, E. R., & Mehdiyeva, G. Y. (2020). Assessment of solar energy potential and its ecological-economic efficiency: Azerbaijan case. Sustainability, 12(3), 1116.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of big data, 7(1), 1-45.

Hemmatabady, H., Welsch, B., Formhals, J., & Sass, I. (2022). AI-based enviro-economic optimization of solar-coupled and standalone geothermal systems for heating https://doi.org/10.3390/su12031116

Huseynli, B. (2023). Renewable Solar Energy Resources Potential and Strategy in Azerbaijan. International Journal of Energy Economics and Policy, 13(1), 31.

Ibrahimov, R. (2010). Azerbaijan energy strategy and the importance of the divercification of exported transport routes. Journal of Qafqaz University, (29).

Khodayar, M., Mohammadi, S., Khodayar, M. E., Wang, J., & Liu, G. (2019). Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. IEEE Transactions on Sustainable Energy, 11(2), 571-583.

Liu, D., & Sun, K. (2019). Random forest solar power forecast based on classification optimization. Energy, 187, 115940.

Mustafayev, F. (2021). The potential role of renewable energy in providing energy security of Azerbaijan. Unpublished. Ph. D. Thesis, University of Gdansk, Sopot, Poland. https://doi.org/10.37474/0365-8554/2020-6-7-57-62

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

Vidadili, N., Suleymanov, E., Bulut, C., & Mahmudlu, C. (2017). Transition to renewable energy and sustainable energy development in Azerbaijan. Renewable and Sustainable Energy Reviews, 80, 1153-1161.

Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons.

Zargar, S. (2021). Introduction to sequence learning models: RNN, LSTM, GRU. Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina, 27606.

https://www.kaggle.com/datasets/ibrahimkiziloklu/solar-radiation-dataset

https://www.kaggle.com/datasets/ishaanthareja007/solar-radiation

https://www.kaggle.com/datasets/pythonafroz/solar-powe-generation-data

https://www.kaggle.com/datasets/vipulgote4/solar-power-generation